

# NEW FORMS OF SPATIAL ANALYSIS WITH MIXED MODEL EFFECTS

by

W. T. Federer, Biometrics Unit and Dept. of Statistical Sciences, Cornell University, Ithaca, N. Y.

J. Crossa, Biometrical Department, CIMMYT, Texcoco, Mexico

J. Franco, Biometrical Department, CIMMYT, Texcoco, Mexico

E. A. Newton, Biometrics Unit and Dept. of Statistical Sciences, Cornell University, Ithaca, N. Y.

April 1998

## 0. ABSTRACT

Analyses for designed spatially laid out experiments need to take account of the design restrictions found in the experiment. What is known about the spatial variation is taken into consideration when selecting the experiment design. The actual spatial variation that occurs during the course of the experiment may not be accounted for by the experiment design or the statistical analysis selected before the experiment was conducted. Hence, to obtain an appropriate statistical analysis for the data obtained from the experiment, it is necessary to determine the type and nature of the spatial variation that is present in the experiment. This often means that an appropriate statistical analysis needs to be selected from a family of plausible statistical analyses. In this paper, standard textbook, polynomial regression (PRTA), principal component (PCTA) and other new forms of spatial analyses are presented. Six different statistical analyses are applied to three data sets. In previous work, it has been found that the regression or trend analysis performed better than nearest neighbor or smoothing for a variety of spatial variation patterns. The new forms of spatial analyses presented herein involve the use of principal components and combinations of principal components and regression procedures. Other forms of spatial analyses are discussed.

### 3NEW FORMS OF SPATIAL ANALYSES WITH MIXED MODEL EFFECTS

#### 1. INTRODUCTION

Published accounts of one form of spatial analysis, i.e., trend analysis (TA), for row-column layouts utilize functions of the row and column effects. The usual form is polynomial regression within complete blocks or across rows and columns. Also, interactions of the row and column regressions may be used to account for non-row-column oriented variation. Such phenomena as disease and insect invasions, lodging patterns, and water flow are common causes of variation not anticipated when the experiment was laid out. When planning and preparing for an experiment, an experiment design (ED) is selected, e. g., a randomized complete block design (RCBD), a Latin square design, an incomplete block design, etc. The selected design is used to lay out the experiment in the field. When analyzing data from the experiment, account needs to be taken of the blocking used as well as how the experiment was actually laid out and what happened during the course of running the experiment. The spatial variation not accounted for by blocking must be taken into account by the statistical analysis. This means that standard textbook analyses for the selected design may be inappropriate in that they may not account for the actual layout and spatial variation present in the experiment.

In the following, spatial analyses are applied to three examples showing considerable spatial variation. The three examples are:

- (i) A randomized complete block designed experiment which was laid out as eight rows and seven columns in the field. The crop was tobacco plants and the response was plant height. The layout and responses are given in Federer and Schlottfeldt (1954).
- (ii) A balanced lattice square designed (BLSD) experiment on counts of cotton plants infected by boll weevil as affected by 16 different insecticides. The data and layout are given in Table 12.5 of Cochran and Cox (1957).
- (iii) A semi-balanced lattice square designed (SBLSD) experiment on yields of ear corn from 25 corn hybrids. The data and layout are presented in Table 12.3 of Cochran and Cox (1957).

Upon investigation, it is shown that the analyses presented by the above authors did not account for the variation present in the experiments.

Six different statistical analyses are applied to these three examples to demonstrate how each controls the heterogeneity present. The six analyses considered are:

- (i) textbook randomized complete block design (RCBD),
- (ii) textbook row-column (RCD) or lattice square (LSD) design,
- (iii) gradients within rows (or columns) of the design (GRADTA),
- (iv) polynomial regressions of rows and columns and interactions (PRTA),
- (v) principle components of rows and columns and interactions (PCTA), and
- (vi) additive row and column effects and multiplicative interactions (AMMITA).

The RCBD and RCD or LSD analyses are discussed in several textbooks and available in several software packages. Federer and Wolfinger (1998) and Wolfinger *et al.* (1997) present codes demonstrating how to perform the computations under a mixed effects model. The GRADTA analysis is an extension of the ideas proposed by Cox (1958). Federer (1998) has described the GRADTA and PRTA analyses, applied them to the balanced lattice square designed experiment described above, and showed how to recover inter-effect information for the adjusted treatment means. Two new forms of spatial analyses, PCTA and AMMITA, will be applied to the two data sets and compared with the other analyses. Other forms and variations of trend analyses are briefly described in the Discussion Section.

A word of caution needs to be noted. When designing an experiment, the *minimum* amount of blocking to control experimental heterogeneity should be used. Over-blocking uses up degrees of freedom and may cause under-estimates of the error mean square. When analyzing data from experiments, care should be taken not to over-parameterize. The *most parsimonious* model that controls the heterogeneity should be used. This maximizes the degrees of freedom associated with the residual mean square.

A description of differential gradients within blocks or rows method (Federer, 1998) is given in the next section. The following section contains a description of polynomial regression trend analysis (PRTA). In the fourth section, a description of principal component analyses is presented. In Section 5, additive main effects and multiplicative interaction models are described. The next three sections contain a description of the six models used for each of the examples and the results obtained. Since the properties of the procedures are unknown, simulation was used to determine appropriate numbers of degrees of freedom to ascribe to the various components of a procedure. A brief discussion of other methods of spatial analyses is given in Section 10.

## 2. GRADIENTS WITHIN BLOCKS (ROWS) ANALYSIS (GRADTA)

If gradients appear within blocks, incomplete or complete, or within rows (columns) of the design and if the gradients are constant or differential, the statistical analysis for an experiment should reflect this. When gradients are present, there is no reason to believe that they are estimates of a common gradient for the entire experiment. Thus, differential gradients rather than a single gradient will be the rule. Also, these gradients should be considered to be random effects since there will usually be no reason to believe that they are estimates of a single parameter. Under this situation, inter-gradient information should be recovered in order to have a more precise estimate of the treatment means. The procedure for doing this has been described by Federer (1998). When there are differential gradients in the complete blocks, the incomplete blocks, or in the rows (columns), GRADTA will account for the spatial variation present. For well-defined patterns, linear and/or quadratic gradients will be sufficient, whereas for spotty or ill-defined variation, higher degree polynomials will be needed. A degree of freedom is used for each regression coefficient in the model. Therefore it is necessary to use as few regressions as possible to account for the variation present. For example, if a single regression is used for each of  $r$  complete blocks (replicates),  $r$  degrees of freedom will be used for each degree of polynomial regression in the model; if differential regressions within each of  $k$  incomplete blocks are used in the model, there will be  $rk$  degrees of freedom used for each polynomial regression. For large incomplete blocks, parsimony in including the additional parameters in the model is not so important but is very important for smaller incomplete blocks as the number of degrees of freedom associated with the residual sum of squares may be less than the number required for the differential gradients.

In the models considered, all effects in the models except for treatments are random effects. The random effect information is used to obtain the adjusted treatment means.

## 3. POLYNOMIAL REGRESSION TREND ANALYSIS (PRTA)

Forms of PRTA have appeared in published literature for many years (e.g., Fisher, 1946; Federer and Schlottfeldt, 1954; Cox, 1958; Federer, Robson, and Tukey, 1962; Tamura *et al.*, 1988). In a row-column layout context, some function, usually polynomial regression, of the row and column effects are used to explain row and column variation. These were treated as fixed effects but may be treated as random effects and this information used to obtain more precise estimates of treatment means as Federer (1998) has shown. For  $r$  rows and  $c$  columns, the question arises as to which of the  $r - 1$  row regressions and the  $c - 1$  column regressions should be used to account for the spatial variation present in an experiment. The same question arises in connection with selecting which interaction terms to include in the model. In selecting the regressions to use in controlling variation, we shall use one of the procedures developed by Bozivich *et al.* (1956) for pooling blocking degrees of freedom and sums of squares with the error or residual degrees of freedom and sum of squares. They have shown that their procedure has little effect on Type I errors. Using available computer software (e.g., Federer and Wolfinger, 1998, Wolfinger *et al.*, 1997), it is a simple task to compute the  $r - 1$  (or fewer) row regressions and the  $c - 1$  (or fewer)

column regressions and their interactions. The following procedure is used to determine which polynomial regressions should be used to explain the spatial variation present in the experiment:

Step 1. Obtain the row regressions,  $R_j$ ,  $j = 1, 2, \dots, r - 1$ , and the column regressions,  $C_i$ ,  $i = 1, 2, \dots, c - 1$ . Compute the sums of squares attributable to each regression. All regressions whose F-values are smaller than the tabulated F-value at the 25% level are relegated to the residual category. The remaining  $C_i$  and  $R_j$  are retained as blocking variables.

Step 2. Determine which interactions  $R_j \cdot C_i$  are to be considered. Suppose that  $i, j = 1, 2, 3, 4$  is the range under consideration. Together with the  $R_j$  and  $C_i$  retained from step 1, compute the sums of squares for the  $R_j \cdot C_i$  interactions. Then use the Bozivich *et al.* (1956) procedure to determine which interactions are to be retained to explain the spatial variation present in the experiment.

Step 3. For the  $R_j$ ,  $C_i$ , and  $R_j \cdot C_i$  regressions retained, compute the analysis using fixed and mixed (random effects) model methods.

Depending upon the size of the experiment, the degree of polynomials under consideration, and the number of interactions under consideration, it may be possible to combine steps 1 and 2. This may be done for the larger experiments as the degrees of freedom associated with the residual sum of squares will be large.

In the models considered, all effects in the models except for treatments are random effects. The random effect information is used to obtain the adjusted treatment means.

#### 4. PRINCIPAL COMPONENT TREND ANALYSIS (PCTA)

As an alternative to polynomial regression trend analysis (PRTA), we propose a new method which is denoted as principal components trend analysis (PCTA). PCTA makes use of the principal components for rows as variables, principal components for columns as variables, and interactions of row and column principal components. The rationale for this approach is that spatial variation in experiments is to be explained as parsimoniously as possible and principal components is one method for doing this. Previous uses of principal components to explain variation is the additive main effects and multiplicative (AMMI) method as described by Gauch (1988) and Zobel (1990, 1992), e.g., and the alternative proposed by Crossa *et al.* (1997) which uses row and column principal components for main effects and multiplicative interactions. These methods were proposed in the context of genotype by environment interaction. For our analyses, the spatial effects will be considered as random effects and mixed model procedures are used. The inter-principal component information is recovered and is used to obtain the adjusted treatment means.

Since principal component analysis obtains linear combinations of the data maximizing sums of squares, it would appear that fewer principal components than polynomial regressions could be used to explain the spatial variation present in an experiment. Both are linear combinations of the data but polynomial regressions are specifically patterned linear combinations. Our procedure is to compute the first few, say three, principal components with rows as variables and then with columns as variables. Interactions of the components are added to account for spatial variation which is other than row-column oriented. Steps 1, 2, and 3 given above are used to determine which row principal components,  $RPC_j$ , which column principal components,  $CPC_i$ , and which interactions  $CPC_i \cdot RPC_j$  to retain in the model.

The above rule for selecting which principal components and which interactions to retain in the spatial model may not be sufficient for PCTA. It appears from the many analyses conducted that the PCTA tends to pick up some of the treatment variation, resulting in lower F-values. In line with the multivariate criterion of maximizing the treatment over the treatment plus error sums of squares, one could use only enough  $CPC_i$ ,  $RPC_j$ , and interactions to obtain a maximum F-value. The result would be to have a parsimonious spatial model which does not diminish the treatment significance. This concept is further discussed in connection with the PCTA for the examples.

## 5. ADDITIVE MAIN EFFECTS AND MULTIPLICATIVE INTERACTION TREND ANALYSIS (AMMITA)

An additive main effects and multiplicative interaction (AMMI) model has received considerable discussion in the literature (See e.g., Gauch, 1988, Zobel, 1990, 1992, and references therein). The idea is to use an additive main effects model for the rows and columns (in our setting) and to use principal components analysis on the residuals from rows, columns, replicates, and treatments. Rather than using linear by linear, linear by quadratic, etc. interactions of row and column regressions as was done for PRTA, we propose to use principal components (PCs) to account for the row by column interactions. The rationale here is that PCs maximize the sum of squares attributable to *any* linear combination of the row-column-replicate-treatment residuals. Thus a PC will account for as much or more of the variation than will any single regression interaction. Fewer PC linear combinations should be needed to explain the spatial variation than required for row by column interactions.

The PCs are random variates just as the regressions were in PRTA as they vary randomly from replicate to replicate or from row to row in a lattice rectangle designed experiment. For a row by column designed experiment, the row effects, the column effects, and the PCs on interaction terms are also considered to be random effects. If one or possibly two PCs control the variation, there is not much information in only one or two degrees of freedom. For the resolvable row-column (lattice rectangle) designs, there will be several degrees of freedom associated with the PCs and inter-principal component information should be recovered as was done for the previous cases in this paper. The method for doing this follows that given by Federer (1998).

In the models considered, all effects in the models except for treatments are random effects. The random effect information is used to obtain the adjusted treatment means.

## 6. EIGHT ROW BY SEVEN COLUMN EXAMPLE

Six analyses are applied to the eight row by seven column designed experiment given in Federer and Schlottfeldt (1954). The models used for the six analyses for the response height are given below in the SAS format. Using these models, SAS GLM and MIXED programs for the six analyses are used as presented in Appendix A.

RCBD: height = replicate treatment

RCD: height = row column treatment

GRADTA: height = row treatment C2\*row C3\*row C4\*row

PRTA: height = treatment C1 C2 C3 C5 R1 R2 R3 R5 R6 R7 C1\*R1 C1\*R2 C2\*R2 C2\*R3 C3\*R2 C4\*R1 C4\*R2

PCTA: height = treatment RPC1 CPC1 CPC2 CPC1\*RPC1 CPC1\*RPC2 CPC1\*RPC3 CPC2\*RPC3 CPC3\*RPC1 CPC3\*RPC2 CPC3\*RPC3

AMMITA: height = column treatment CPC1\*column CPC2\*column

C<sub>i</sub> refers to the *i*th column polynomial regression and R<sub>j</sub> is the *j*th row polynomial regression, *i* = 1, 2, ..., *c* - 1, and *j* = 1, 2, ..., *r* - 1. Note that not all regressions may be computed. CPC<sub>i</sub> is the *i*th column principal component and RPC<sub>j</sub> is the *j*th row principal component. In SAS methodology, the asterisk is used for interactions and for nesting a covariate (a continuous variable) within a class (discrete) variable. The SAS statement of a model omits the subscripts, the plus signs, and the error term. For example a linear model for a row-column design is

$$Y_{hij} = \mu + \rho_h + \gamma_i + \tau_j + \varepsilon_{hij}$$

where  $\mu$  is a general mean effect,  $\rho_h$  is the  $h$ th row effect,  $\gamma_i$  is the  $i$ th column effect,  $\tau_j$  is the  $j$ th treatment effects, and  $\varepsilon_{hij}$  is a random error term. The SAS model is written as above for the RCD.

In Table 1, the residual degrees of freedom and mean square, the treatment mean square, the F-value for the treatments, and coefficient of variation are presented for the six analyses. The coefficient of variation for the RCBD analysis is 17%. There is a large column effect which inflates the RCBD residual to 30,228 from the row-column residual mean square of 7,352 with a coefficient of variation of 8%, less than half that of the RCBD analysis. The F-value for treatments is nearly doubled, increasing from 1.51 to 2.71. Note the large decrease in the treatment mean square for the RCD analysis.

The differential gradient within block analysis, GRADTA, proposed by Federer (1998) is presented in Table 1. There appeared to be no discernable linear gradients within blocks but the C2\*row (quadratic within row), C3\*row (cubic within row), and C4\*row (quartic within row) gradients all appeared to be present. 24 degrees of freedom were used to account for the differential gradients, leaving only 18 for the residual mean square. If the 24 degrees of freedom were partitioned into single degree of freedom contrasts, several of these would be pooled with the residual, thus possibly reducing the residual mean square from the obtained value of 11,310. This analysis did not result in much of a decrease of the treatment mean square but did result in a larger F value, 3.78, than the above analyses. Instead of considering differential gradients in rows as above, differential gradients within columns may be considered. Using the model

$$\text{height} = \text{column treatment } R1 * \text{column } R2 * \text{column.}$$

The residual mean square for this model was 6,980, and the F value was 1.98.

When the above PRTA analysis is applied to the data set and using Step 1 from above, the polynomial regressions C6 and R4 are the only ones relegated to residual. Since such high ordered regressions are retained, no simple pattern of variation exists. Proceeding to Step 2 and the interactions  $C_i * R_j$ ,  $i, j = 1, 2, 3, 4$ , the interactions retained are C1\*R1, C2\*R2, C2\*R3, C3\*R2, C4\*R1, and C4\*R2. The other 10 interaction terms and C4 had F-values less than F at the 25% level. Note that there may be higher degree polynomial interactions than the 16 considered that could possibly be retained in the model. The less organized the spatial variation patterns, the higher will the degree of the polynomial regressions retained. The resulting residual and treatment mean squares are given in Table 1. The PRTA analysis showed considerable reduction in the residual mean square over the row-column analysis, i.e., 4,204 versus 7,352. Both of these mean squares are considerably smaller than the RCBD mean residual square of 30,228. The significance of the treatment mean square should also be noted as the RCBD analysis showed no significance at the 5% level, say, whereas the F-value of 6.36 for PRTA was considerably larger than for the RCBD, RCD, or GRADTA analyses. Combining Steps 1 and 2 results in the same model and may be used here as there are sufficient degrees of freedom to account for all regressions investigated and still leave 20 or more for the residual mean square.

CPC<sub>i</sub> and RPC<sub>j</sub> principal components were computed for  $i, j = 1, 2, 3$ . The nine interactions CPC<sub>i</sub>\*RPC<sub>j</sub> were also computed for the data set of Federer and Schlottfeldt (1954). Applying Steps 1, 2, and 3, RPC1, CPC1, CPC2, and the seven interactions CPC1\*RPC1, CPC1\*RPC2, CPC1\*RPC3, CPC2\*RPC3, CPC3\*RPC1, CPC3\*RPC2, and CPC3\*RPC3 were retained in the model. RPC2, RPC3, CPC3, CPC2\*RPC1, and CPC2\*RPC2 were relegated to the residual category. As may be observed from Table 1, there is a dramatic reduction in the residual mean square, i.e., 2,498. The residual mean square obtained is 1/12th of the RCBD mean square of 30,228, 1/3rd of the row-column mean square of 7,352, and 3/5th of the PRTA residual mean square of 4,204. This PCTA analysis reduced the coefficient of variation to 5% from 17% for the RCBD analysis. The p-values for the F statistics for treatment effect were 20% for the RCBD analysis, 3% for the row-column analysis, 1% for differential gradients, less than 1% for the PRTA, and 5% for the PCTA. Although the residual mean square was reduced, so was the treatment mean square, being only about one-ninth of that for the RCBD and about one-fifth of the treatment mean square from PRTA. This resulted in a smaller F value.

The above PCTA model appears to control the variation in the experiment very well but at the expense of reducing the variation among treatments. The PCTAp model which is parsimonious, controls the variation well, and yields a larger F-value, is

height = treatment CPC1 CPC2 RPC1 CPC1\*RPC1 CPC1\*RPC2.

This model results in a residual mean square of 4,726, nearly as small as PRTA and an F-value of 2.97 versus 2.32 for the above less parsimonious model. In Table 1, the preceding model is denoted as PCTAp and the larger model as PCTA.

The AMMITA model resulted in a residual mean square of 4,744 and an F value of 3.25. PRTA, PCTAp, and AMMITA analyses resulted in approximately equal residual mean squares and coefficients of variation. The spatial variation controlled by these three procedures appears to be sufficient. The PCTA model appears to be an over parameterization and to have taken up some of the variation among treatments. The treatments adjusted for the random effects in the model are given in Table 2. The means are ranked from highest to lowest according to those obtained from the PRTA analysis. The results are also presented graphically in Figure 1. The largest change in rankings is for treatments A and C for the GRADTA method.

For this data set, the PRTA model would be the preferred one as the properties of the model are known and it controls the variation present in the experiment.

## 7. BALANCED LATTICE SQUARE EXAMPLE

As a second example, consider the balanced lattice square designed experiment given in Table 12.5 of Cochran and Cox (1957). SAS PROC GLM and MIXED programs for the six models are given in Appendix B. The six models used for explaining the variation present are:

RCBD: count = replicate treatment

BLSD: count = replicate treatment row(replicate) column(replicate)

GRADTA: count = replicate treatment row(replicate) C1\*row(replicate)

PRTA: count = replicate treatment R1\*replicate R2\*replicate C1\*replicate C1\*R1\*replicate C2\*R1\*replicate C2\*R2\*replicate C3\*R2\*replicate

PCTA: count = replicate treatment CPC1\*replicate CPC2\*replicate RPC1\*replicate CPC1\*RPC1\*replicate CPC1\*RPC2\*replicate CPC2\*RPC1\*replicate CPC2\*RPC2\*replicate

AMMITA: count = replicate treatment row(replicate) RPC1\*row(replicate)

Row and CPCi are completely confounded as is column and RPCj.

Before considering the results from the six analyses, some details of the experiment are in order. The response variable is the mean of three separate counts of boll weevils on cotton plants. If count is a Poisson variable, the error mean square should be the mean of the experiment divided by three, or  $10.91/3 = 3.64$ , which is approximately 4. The expected coefficient of variation then would be  $(3.64)^{0.5}/10.91 = 17.5\%$ . However, since the counts were made at three successive times on the same plots, they are highly correlated and this would mean that the expected variance should be somewhat less than the mean 10.91 but higher than 3.64. Also, there were 16 insecticide treatments used and it is highly unlikely that the experimenter would be conducting such an experiment unless treatment differences were expected, i.e., the null hypothesis is not true and significance among the treatments is expected. However, it is to be noted that non-significance could arise when there are no or small infestations of boll weevils. One further item to note is that the experiment was laid out in 20 rows and 4 columns as depicted in Table 12.5 of Cochran and Cox (1957). It is not likely that an insect invasion in the experiment would be in the row-column orientation used for the experiment. Even if count is a Poisson variable, other sources of

error enter as well meaning that the error mean square should be larger than 3.64. Transformation of the count data (log, square root, arcsine) was attempted but the results indicated that untransformed counts should be used (Federer, 1998). This is expected since the data are means of three counts rather than counts.

Even though a coefficient of variation of 57%, Table 3, was obtained for the RCBD analysis, the treatment differences were significant at the 5% level. The standard textbook analysis for a lattice square designed experiment reduced the RBCD residual mean square of 38.88 to 22.67. The coefficient of variation was reduced from 57% to 44%. The resulting F-value of 0.94 is not what would be expected for this type of experiment. Thus the textbook analysis for a lattice square design appears not to account for the expectations for this experiment in several respects.

Applying the GRADTA to the data set (Federer, 1998) using the above model resulted in a residual mean square of 18.97, a coefficient of variation of 40% and an F-value of 1.22. This analysis was more in line with expectations than the standard textbook analysis but still appears to have more spatial variation that needs to be taken into account

The PRTA model described above resulted in a residual mean square of 8.95, an F-value of 3.14, and a coefficient of variation of 27%. For the PRTA used by Federer (1998), only linear and quadratic regressions for rows, columns, and their interactions were used, and the reduction in the obtained residual mean square of 11.91 produced a reduction of 47% over that from the textbook analysis. When the cubic regressions and interactions are included and those with F-values of less than the tabulated F-value at the 25% point were relegated to residual, the further reduction to 8.95 was obtained. This mean square is 39% of the textbook residual, or a reduction of 61%. The model was obtained using Rj\*replicate, Ci\*replicate, and Ci\*Rj\*replicate for i, j = 1, 2, 3. C2\*replicate, C3\*replicate, R3\*replicate, C1\*R2\*replicate, C1\*R3\*replicate, C2\*R2\*replicate, C2\*R3\*replicate, C3\*R1\*replicate, and C3\*R3\*replicate F-values were all less than the tabulated F-value at the 25% level and hence these regressions were relegated to the residual sum of squares by the above procedure of selection. If there is extra-Poisson variation in this experiment, which could arise from several sources, then the results from Poisson variations do not hold and one may decide not to search further for an appropriate analysis for this data set.

Since there are only four rows and columns in each block of the lattice square designed experiment in Table 12.5 of Cochran and Cox (1957), only CPC1, CPC2, RPC1, RPC2, and their interactions were considered for the PCTA procedure. All of the components were retained in Step 3 except for RPC2\*replicate. To illustrate the effects of this analysis, the Type III sums of squares accounted for by some regression and principal components are

R1*rep	809.9	CPC1*rep	1145.2
R2*rep	177.0	CPC2*rep	56.2
C1*rep	352.1	RPC1*rep	401.4
C1*R1*rep	96.3	CPC1*RPC1*rep	197.1
C2*R1*rep	204.4	CPC1*RPC2*rep	68.0
Total	1639.7	Total	1867.9

As expected, these principal components and two of their interactions were associated with a larger sum of squares than were the three regressions and the linear by linear and quadratic by linear interactions. It appears that CPC1\*rep and CPC2\*rep should be compared with R1\*rep and R2\*rep, and that RPC1\*rep should be compared with C1\*rep sums of squares, and this accounts for the alignment above.

The selection of an appropriate PCTA model poses difficulties as demonstrated using the following models for count:

model 1:

```
count = REP TREAT CPC1*REP CPC2*REP RPC1*REP CPC1*RPC1*REP CPC1*RPC2*REP
CPC2*RPC1*REP CPC2*RPC2*REP
```

model 2:

```
count = REP TREAT CPC1*REP RPC1*REP CPC1*RPC1*REP CPC1*RPC2*REP
CPC2*RPC1*REP CPC2*RPC2*REP
```



```

model 3:
count = REP TREAT CPC1*REP RPC1*REP CPC1*RPC1*REP CPC2*RPC1*REP
CPC2*RPC2*REP
model 4:
count = TREAT RPC1*ROW*REP CPC1
model 5:
count = TREAT CPC1*REP RPC1*REP CPC1*RPC1*REP
model 6:
count = REP TREAT CPC1 CPC1*RPC1*REP
model 7:
count = REP TREAT CPC1*REP CPC2*REP RPC1*REP RPC2*REP

```

Some results from these seven models are:

	degrees of freedom	residual mean square	F-value for treatments
model 1	25	4.22	1.33
model 2	30	5.39	1.26
model 3	35	6.41	0.96
model 4	43	11.66	1.20
model 5	49	12.69	0.95
model 6:	53	13.25	1.00
model 7	40	18.73	1.02

If Poisson variation is the expected variation, then models 1, 2, and 3 are coming close to achieving the expected value of 3.64. Although at least 25 degrees of freedom are associated with the residual mean squares, there could be some over-parameterization in these models.

The treatment means adjusted for random inter-effect information are given in Table 4 ranked according to the PRTA means. All effects except treatment effects in the above six models are considered to be random effects for the models used in SAS PROC MIXED.

## 8. SEMI-BALLANCED LATTICE SQUARE EXAMPLE

The data set given in Table 12.3 of Cochran and Cox (1957) is used for an example of a semi-balanced lattice square design (SBLSD). 25 corn hybrids were arranged in five rows and five columns within each of three replicates. The experimental unit or plot size was four rows by ten hills per row. The response is weight of ear corn in pounds. The response models used are:

RCBD: weight = replicate treatment

BSLSD: weight = replicate treatment row(replicate) column(replicate)

GRADTA: weight = replicate treatment row(replicate) C1\*row(replicate)

PRTA: weight = replicate treatment C1\*replicate R1\*replicate R2\*replicate R3\*replicate  
R4\*replicate C1\*R1\*replicate C1\*R2\*replicate

PCTA: weight = replicate treatment CPC1\*replicate CPC2\*replicate RPC1\*replicate  
RPC2\*replicate CPC1\*RPC1\*replicate CPC1\*RPC2\*replicate CPC2\*RPC1\*replicate  
CPC2\*RPC2\*replicate

AMMITA: weight = replicate treatment row(replicate) RPC1\*row(replicate)

The symbols used in the above models have been defined in previous sections.

The results obtained for the six models are given in Table 5 for degrees of freedom in the residual mean square, the residual mean square, the treatment or hybrid mean square, the treatment F value, and the coefficient of variation. As indicated by the RCBD analysis, the coefficient of variation is rather high for most corn experiments and there is no indication of a differences among the 25 corn hybrids. These are probably indications that more of the spatial variation needs to be taken into account. The textbook analysis for a semi-balanced lattice square as given by Cochran and Cox (1957) has a reasonable coefficient of variation and the treatment mean square is almost significant at the 5% level (1.98). Upon running the GRADTA model, it is apparent that there are differential gradients within the rows which the column effects do not take into account. The GRADTA would appear to be the appropriate analysis for this data set. A coefficient of variation of 6.9% and a treatment F-value of 3.64 is obtained indicating differences among the 25 hybrids, perhaps an anticipated result. The residual mean square is only 43% of the SBLSD mean square of 9.57, i.e., more than doubling the number of replications for the SBLSD analysis. The PRTA, PCTA, and AMMITA analyses did not account for as much of the spatial variation as did the GRADTA analysis.

The mixed model means for the six analyses are given in Table 6. The hybrids are ranked according to the GRADTA analysis means. The top hybrid, 19, was ranked number one by all models except RCBD. The lowest two ranked means, 24 and 25, by all models were common in the lowest three ranked means. There was variation in the rankings by the various methods for the remaining means.

## 9. SIMULATION RESULTS

Although a principal component is nothing but a linear combination of the data in a similar manner as polynomial regression, it is not clear that one degree of freedom should be allocated to each principal component or to each interaction of two components. Therefore, a simulation study for an eight row by seven column layout was used. Random unit normal deviates were used. Doing this, all mean squares in a statistical analysis should equal one within sampling errors. Also, the sum of squares for any effect gives the degrees of freedom associated with that effect. In this manner, it is possible to approximate the degrees of freedom for each of the principal components and for their interactions (Newton, 1997). The results of 1,000 simulations for the eight row by seven column design both with zero and non-zero treatment effects. The latter example is included to measure possible biases related to the analyses.

## 10. DISCUSSION

One of the new forms of spatial analyses proposed for a row-column or lattice rectangle design is to perform an AMMITA analysis on the row  $\times$  column interaction effects. These analyses probably will only be useful for relatively large row-column or lattice rectangle designs and for irregular spatial patterns. Large lattice designs in a row-column arrangement within complete blocks or large resolvable row-column designs, latinized or not, are frequently used for plant breeding experiments. These designs are candidates for the types of analyses discussed herein.

Another form of spatial analysis would be to combine PRTA and PCTA in the manner of AMMI analyses which would take account of the polynomial regressions for row and column effects and principal components for the row-column interactions. This method is denoted as PRAMMITA and the regression effects and the principal component effects would be considered to be random effects. Inter-regression and inter-AMMI information would be recovered to adjust the treatment means.

Still another form of spatial analysis of the trend analysis nature is to use a Fourier analysis in place of polynomial regression trend analysis, i.e., FRTA. Such an analysis as a FRTA will be useful when the spatial variation can be explained by trigonometric functions such as sine and cosine, i.e., there is some form of cyclical patterns in the variation. Perhaps variation caused by damage to one spot in the experiment can be accounted for by a Fourier analysis. Alternatively, the damaged area may be denoted as another block or the damage may be measured or graded and used as a covariate.

Another form of spatial analysis would be to use a multiplicative interaction procedure along with the PCTA analysis on rows and columns. That is, the residuals after fitting PCs on rows and

columns are obtained and then a PC analysis is performed on these residuals. This analysis would differ from the PCTA method of using  $CPCi*RPCj$  interactions.

Two different programs are presented for an AMMITA analysis. One is contained in the program for PCTA analyses, and the other is presented for AMMITA alone. The two programs do not yield the same results. The results for the row-column design for the residual mean square are 4744 versus 4796 when using  $CPC1*col$  and  $CPC2*col$  from the PCTA analysis instead of  $prin1*col$  and  $prin2*col$  from the AMMITA program. Although this difference is small here, this was not the case for the other two examples. A study needs to be made to ascertain the nature of the difference in the two programs. For the present it is recommended that  $CPCi*col$  be used in place of  $prini*col$ .

Model	Degrees of freedom	Residual mean square	Treatment		Coefficient of variation
			mean square	F-value	
RCBD	42	30,228	45,646	1.51	17
RCD	36	7,352	19,954	2.71	8
GRADTA	18	11,310	42,730	3.78	11
PRTA	33	4,204	26,743	6.36	6
PCTAp	44	4,726	14,047	2.97	7
PCTA	39	2,498	5,787	2.32	5
AMMITA	29	4,744	15,417	3.25	7

Table 1. Results of six analyses for eight row by seven column data set.

Model	Treatment						
	C	D	B	A	F	E	G
RCBD	1043	999	1084	1059	1043	975	864
RCD	1051	1030	1054	1031	1025	996	899
GRADTA	1123	1006	1054	1098	995	945	867
PRTA	1083	1063	1036	1031	1019	955	899
PCTAp	1044	1039	1045	1020	1017	1005	917
PCTA	1042	1020	1039	996	1031	1000	957
AMMITA	1058	1012	1058	1043	1017	989	910

Table 2. Mixed model plant height means (recovery of inter-effect information) for the six analyses for eight row by seven column data set. Treatments ranked by PRTA means.

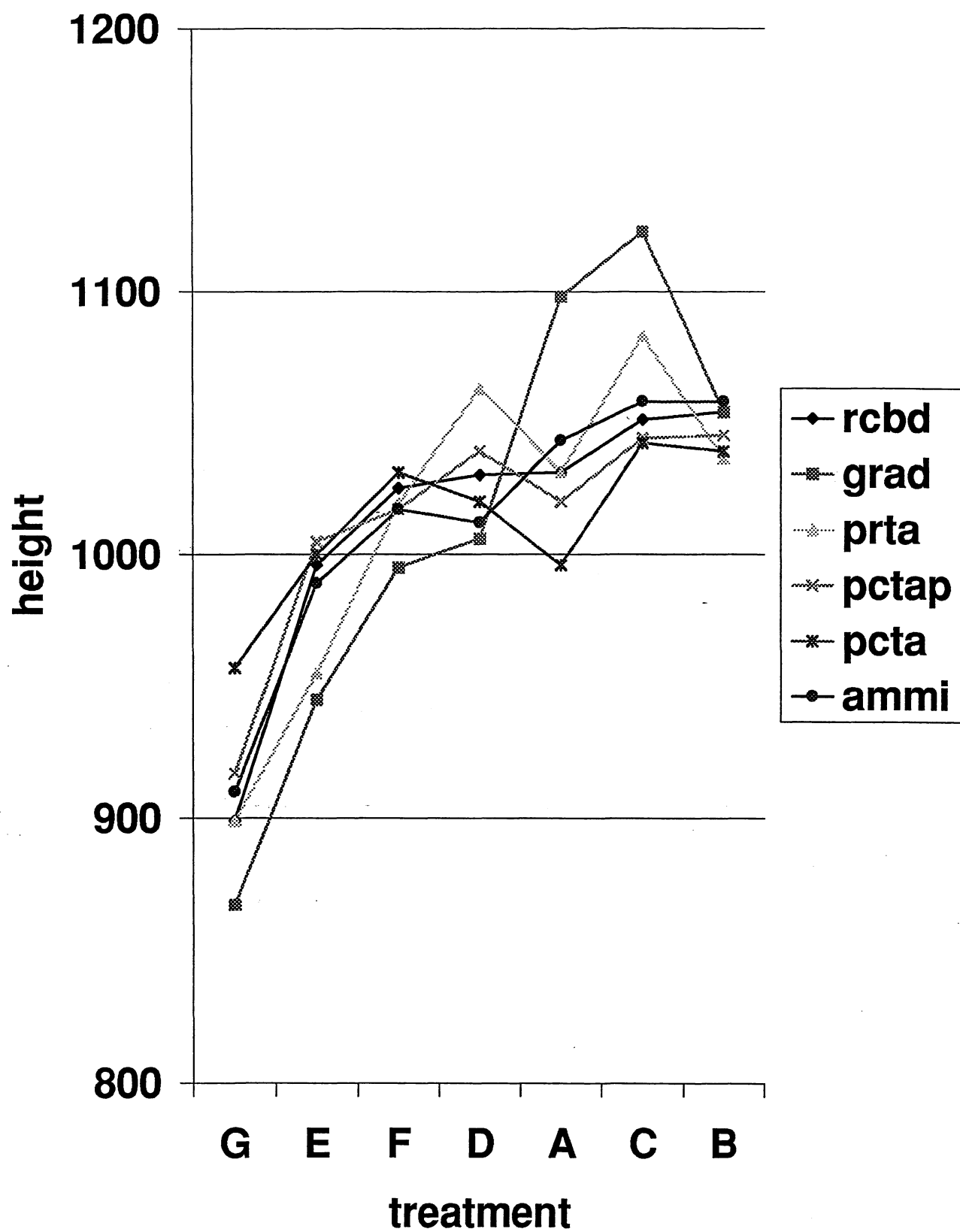


Figure 1. Graphical representation of mixed effect means for six models for the eight row by seven column data set.

Model	Degrees of freedom	Residual mean square	Treatment mean square	F-value	Coefficient of variation %
RCBD	60	38.88	82.95	2.13	57
BLSD	30	22.67	21.30	0.94	44
GRADTA	25	18.97	23.15	1.22	40
PRTA	25	8.95	28.10	3.14	27
PCTA	25	4.22	5.60	1.33	19
AMMITA	25	11.10	15.28	1.38	31

Table 3. Results of six analyses for balanced lattice square data set.

Treatment number	RCBD	LSD	GRADTA	PRTA	PCTA	AMMITA
11	19.6(1)	17.9(1)	17.7(1)	15.3(1)	13.5(1)	15.3(1)
14	14.9(3)	14.3(3)	13.2(5)	15.2(2)	11.8(4)	13.9(3)
10	16.3(2)	15.2(2)	15.4(2)	15.2(2)	12.8(2)	14.7(2)
16	8.3(12)	10.6(8)	10.0(9)	13.7(4)	10.6(10)	12.4(5)
2	14.9(4)	13.7(4)	15.0(3)	13.4(5)	8.8(16)	10.6(10)
5	10.0(8)	9.6(10)	9.4(10)	12.0(6)	12.2(3)	10.8(9)
4	11.2(7)	11.3(6)	12.1(6)	11.6(7)	11.1(7)	12.3(7)
12	13.7(5)	12.8(5)	13.4(4)	11.5(8)	9.8(13)	13.6(4)
8	9.7(10)	9.4(11)	9.4(11)	11.4(9)	11.3(6)	10.2(11)
13	12.2(6)	11.0(7)	10.7(7)	10.4(10)	10.9(9)	11.5(8)
9	10.0(9)	10.1(9)	9.2(12)	9.9(11)	10.9(8)	12.3(6)
3	6.7(14)	8.5(13)	8.8(13)	9.8(12)	9.6(15)	7.9(14)
15	8.7(11)	9.3(12)	10.2(8)	8.5(13)	11.3(5)	8.6(12)
6	6.8(13)	7.3(14)	8.3(14)	6.5(14)	10.1(11)	8.2(13)
1	4.9(16)	6.1(16)	5.5(16)	5.1(15)	9.7(14)	6.4(15)
7	6.7(15)	7.3(15)	6.2(15)	4.9(16)	10.0(12)	5.7(16)

Table 4. Mixed model count means (recovery of inter-effect information) for the six analyses for the balanced lattice square example. Treatments ranked (ranks in parentheses) according to PRTA means.

Model	Degrees of freedom	Residual mean square	Treatment mean square	F-value	Coefficient of variation %
RCBD	48	21.95	25.46	1.16	16.0
SBLSD	24	9.57	18.10	1.89	10.6
GRADTA	21	4.06	14.78	3.64	6.9
PRTA	27	6.66	14.96	2.25	8.8
PCTA	24	5.00	7.97	1.59	7.7
AMMITA	21	5.87	11.96	1.97	8.3

Table 5. Results of six analyses for semi-balanced lattice square data set.

Hybrid	RCBD	SBLSD	GRADTA	PRTA	PCTA	AMMITA
19	34.2(2)	35.7(1)	35.5(1)	35.0(1)	34.2(1)	35.3(1)
22	29.5(15)	31.5(6)	33.1(2)	32.5(2)	31.2(4)	31.3(4)
18	29.9(13)	31.3(7)	32.9(3)	32.3(4)	29.8(8)	31.0(6)
11	27.4(19)	30.0(8)	32.4(4)	31.0(7)	31.2(4)	31.0(6)
15	31.4(6)	32.9(3)	32.1(5)	32.5(2)	30.3(7)	31.7(3)
8	34.5(1)	33.4(2)	31.8(6)	31.6(5)	32.5(2)	32.1(2)
10	31.7(5)	32.5(4)	30.3(7)	31.5(6)	32.0(3)	30.7(8)
7	29.4(16)	27.2(19)	30.1(8)	29.4(11)	29.7(9)	28.2(15)
5	34.1(3)	31.8(5)	30.1(8)	30.6(9)	30.6(6)	31.2(5)
12	30.5(9)	30.0(8)	29.8(10)	31.0(7)	27.6(22)	29.4(11)
6	27.5(18)	28.1(16)	29.4(11)	28.3(16)	29.2(12)	29.5(10)
2	29.6(14)	28.6(14)	29.4(11)	28.5(13)	28.2(18)	30.3(9)
4	30.1(11)	29.6(10)	28.8(13)	28.7(12)	29.3(11)	29.4(11)
9	31.1(8)	29.1(12)	28.0(14)	28.5(13)	29.6(10)	29.1(14)
14	27.7(17)	28.7(13)	28.7(15)	28.4(15)	28.4(15)	28.1(16)
13	26.2(21)	26.2(22)	28.3(16)	26.6(22)	26.4(23)	27.2(21)
17	30.1(11)	27.9(17)	28.2(17)	27.9(18)	28.3(16)	29.4(11)
16	31.3(7)	28.5(15)	27.4(18)	28.1(17)	29.1(13)	28.0(18)
3	26.3(20)	29.3(11)	27.3(19)	29.8(10)	28.3(17)	27.4(20)
24	25.8(22)	26.6(21)	26.5(20)	27.2(20)	27.7(21)	28.1(16)
1	26.2(21)	27.9(17)	26.0(21)	27.3(19)	28.1(19)	27.5(19)
21	25.7(23)	26.0(23)	25.8(22)	26.1(23)	28.5(14)	26.3(23)
23	24.8(24)	24.2(25)	25.8(22)	25.1(25)	26.1(24)	24.5(25)
25	30.4(10)	27.2(19)	25.7(24)	26.8(21)	25.9(25)	26.3(23)
20	24.3(25)	25.7(24)	25.4(25)	25.2(24)	27.8(20)	26.7(22)

Table 6. Mixed model means (recovery of inter-effect information) in pounds of ear corn for the six analyses for the semi-balanced lattice square example. Treatments ranked (ranks in parentheses) according to GRADTA means.

## 11. LITERATURE CITED

- Bozivich, H., T. A. Bancroft, and H. O. Hartley (1956). Power of analysis of variance test procedures for certain incompletely specified models. *Annals of Mathematical Statistics* 27:1017-1043.
- Cochran, W. G. and G. M. Cox (1957). *Experimental Designs*, second edition. John Wiley & Sons, Inc. New York.
- Cox, C. P. (1958). The analysis of Latin square designs with individual curvatures in one direction. *J. Royal Statistical Soc., Series B*, 20:193-204.
- Crossa, J. (1997).
- Federer, W. T. (1998). Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. *Biometrics* 54(2): (in press)
- Federer, W. T., E. A. Newton, and N. S. Altman (1997). Combining standard block analyses with spatial analyses under a random effects model. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications, and Future Directions* (Eds: T. G. Gregoire *et al.*), Springer, New York, Heidelberg, pp. 373-386.
- Federer, W. T., D. S. Robson, and J. W. Tukey (1962). Tests for non-additivity. BU-147-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, N.Y.
- Federer, W. T. and C. S. Schlottfeldt (1954). The use of covariance to control gradients in experiments. *Biometrics* 10:282-290.
- Federer, W. T. and R. D. Wolfinger (1998). SAS PROC GLM and PROC MIXED code for recovering inter-effect information. *Agronomy J.* 90:(in press)
- Fisher, R. A. (1946). *Statistical Methods for Research Workers*. Oliver & Boyd, Ltd, Edinburgh
- Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics* 44:705-715.
- Newton, E. A. C. (1997). Accounting for spatial variation: A comparison of methods. Unpublished Paper, Biometrics Unit, Cornell University.
- Tamura, R., L. Nelson, and G. Naderman (1988). An investigation of the validity and usefulness of treend analysis. *Agronomy J.* 80:712-718.
- Wolfinger, R. D. W. T. Federer, and O. Cordero-Brana (1997). SAS PROC GLM and MIXED for recovering information in augmented designs. *Agronomy J.* 89(6):856-859.
- Zobel, R. W. (1990). A powerful statistical model for understanding genotype-by-environment interaction. In *Genotype-by-Environment Interaction and Plant Breeding* (Ed: M. S. Kang). Louisiana State Univ., Baton Rouge, pp. 126-140.
- Zobel, R. W. (1992). Soil environment constraints to root growth. In *Advances in Soil Science* (ed: J. L. Hatfield and B. A. Stewart). Springer-Verlag, New York, 19:27-51.



## Appendix A. Programs for six analyses for eight row by seven column data set.

Let col = column and treat = treatment for the following programs for the file named colrow.dat:

### RCBD

```
data colrow;
  infile 'colrow.dat';
  input yield row col treat;
proc glm data = colrow;
  class row col treat;
  model yield = row treat; /*Add "/solution" after "treat" for effects and standard errors of a
difference.*/
  lsmeans treat;
run;
```

### RCD

```
data colrow;
  infile colrow.dat';
  input yield row col treat;
proc glm data = colrow;
  class row col treat;
  model yield = row col treat; /*Add "lsmeans treat;" for fixed effect means.*/
run;
proc mixed data = colrow;
  class row col treat;
  model yield = treat;
  random row col;
  lsmeans treat;
run;
```

### GRADTA and PRTA

```
proc iml;
  opn7 = orpol(1:7,6); /*"7,6" means that there are 7 columns and polynomials up to 6th degree.*/
  opn[7,1] = (1:7)';
  op7 = opn7;
  create opn7 from opn7[colname = {'COL' 'C1' 'C2' 'C3' 'C4' 'C5' 'C6'}];
  append from opn7;
  close opn7;
run;
  opn8 = orpol(1:8,7); /*There are 8 rows.*/
  opn8[1,1] = (1:8)';
  op8 = opn8;
  create opn8 from opn8[colname = {'ROW' 'R1' 'R2' 'R3' 'R4' 'R5' 'R6' 'R7'}];
  append from opn8;
  close opn8; run;
data colrow;
  infile 'colrow.dat';
  input yield row col treat;
data crbig; set colrow;
  idx = _n_; run;
proc sort data = crbig;
```

```

    by COL; run;
data crbig;
    merge crbig opn7;
    by COL; run;
proc sort data = crbig;
    by ROW; run;
data crbig;
    merge crbig opn8;
    by ROW; run;
proc sort data = crbig;
    by idx; run;
proc glm data = crbig;
    class row col treat;
    model yield = row treat C2*row C3*row C4*row; /*Final model.*/
run;
proc mixed data = crbig;
    class row col treat;
    model yield = treat;
    random row C2*row C3*row C4*row;
    lsmeans treat;
run;
proc glm data = crbig;
    class row col treat;
    model yield = treat C1 C2 C3 C5 R1 R2 R3 R5 R6 R7 C1*R1 C2*R1 C2*R3 C3*R2 C4*R1
    C4*R2; /*Final model.*/
run;
proc mixed data = crbig;
    class row col treat;
    model yield = treat;
    random C1 C2 C3 C5 R1 R2 R3 R5 R6 R7 C1*R1 C2*R1 C2*R3 C3*R2 C4*R1 C4*R2;
    lsmeans treat;
run;

```

#### PCTA and AMMITA

```

data colrow1; infile 'colrow1.dat'; input Y row col treat rep;
label rows = 'incomplete block'; /*This data set has rep = 1 for all observations.*/
proc sort data = colrow1; by REP ROW COL;
proc iml; use colrow1;
    READ ALL VAR{REP ROW COL TREAT} INTO CLASS;
    NREP = CLASS[<>,1];
    NRO = CLASS[<>,2];
    NCO = CLASS[<>,3];
    NTR = CLASS[<>,4];
    N = NROW(CLASS);
    A2 = J(NREP*NRO,NCO+2,0);
    CONT1 = 0;
    ROWS = 1:NRO;
    DO I = 1 to NREP;
        READ ALL WHERE(REP = I) VAR(Y) INTO A0;
        A1 = SHAPE(A0,NRO,NCO);
        A2[CONT1+1:CONT1+NRO,] = REPEAT(I,NRO,1) || ROW || A1;
        CONT1 = CONT1 + NRO;
    END;
    NAMES0 = 'COL1':'COL7'; NAMES = {REP ROW} || NAMES0;

```

```

CREATE SPAT FROM A2 [COLNAMES = NAMES];
APPEND FROM A2; QUIT;
PROC PRINT DATA = SPAT; RUN;
DATA COLROW; SET SPAT; IF REP NE 1 THEN DELETE;
PROC PRINCOMP DATA = COLROW PREFIX = CPC OUT = COLVAR NOPRINT;
VAR COL1-COL7;
PROC PRINT DATA = COLVAR;
DATA T1; SET COLVAR; DROP REP ROW;
PROC TRANSPOSE DATA = T1 OUT = ROWCOL PREFIX = ROW;
PROC PRINCOMP DATA = ROWCOL PREFIX = RPC OUT = T2 NOPRINT;
VAR ROW1-ROW8;
DATA ROWVAR; SET T2; COL = _NAME_; DROP _NAME_;
PROC PRINT DATA = ROWVAR;
  data c1; set colrow; y = cOL1; col = 'COL1'; keep row col y;
  data c2; set colrow; y = cOL2; col = 'COL2'; keep row col y;
  data c3; set colrow; y = cOL3; col = 'COL3'; keep row col y;
  data c4; set colrow; y = cOL4; col = 'COL4'; keep row col y;
  data c5; set colrow; y = cOL5; col = 'COL5'; keep row col y;
  data c6; set colrow; y = cOL6; col = 'COL6'; keep row col y;
  data c7; set colrow; y = cOL7; col = 'COL7'; keep row col y;
proc sort; by row col; run;
  DATA CC; SET COLVAR; DROP COL1-COL7 CPC4-CPC7; PROC PRINT;
  DATA RR; SET ROWVAR; DROP ROW1-ROW8 RPC4-RPC8; PROC PRINT;
PROC SORT DATA = T3; BY ROW;
  DATA T6; MERGE T3 CC; BY ROW;
  PROC SORT DATA = T6; BY COL;
  DATA TODO0; MERGE T6 RR; BY COL;
  DATA TODO1; SET TODO0; COL = 1*COMPRESS(COL,'COL');
  PROC SORT; BY ROW COL;
  DATA CC; SET COLVAR; DROP COL1-COL7 CPC4-CPC7; PROC PRINT;
  DATA RR; SET ROWVAR; DROP ROW1-ROW8 RPC4-RPC8; PROC PRINT;
  PROC SORT DATA = T3; BY ROW;
  PROC SORT DATA = T6; BY COL;
  DATA TODO0; MERGE T6 RR; BY COL;
  DATA TODO1; SET TODO0; COL = 1*COMPRESS(COL,'COL');
  PROC SORT; BY ROW COL;
  DATA TODO; SET TODO1;
  PROC SORT DATA = COLROW1 OUT = T1; BY REP ROW COL;
  DATA T2; SET T1; KEEP REP ROW COL TREAT;
  DATA T3; SET TODO; C = 1*COL; DROP COL;
  DATA T4; SET T3; COL = C; DROP C;
  DATA ANA; MERGE T2 T4; BY REP ROW COL;
proc glm data = ANA;
  class row col treat;
  model Y = treat CPC1 CPC2 RPC1 CPC1*RPC1 CPC1*RPC2 CPC1*RPC3 CPC2*RPC3
          CPC3*RPC1 CPC3*RPC2 CPC3*RPC3; /*Final model.*/
run;
proc mixed data = ANA;
  class row col treat;
  model Y = treat;
  random CPC1 CPC2 RPC1 CPC1*RPC1 CPC1*RPC2 CPC1*RPC3 CPC2*RPC3 CPC3*RPC1
          CPC3*RPC2 CPC3*RPC3;
  lsmeans treat;
run;

```

```

proc glm data = ANA;
  class row col treat;
  model Y = col treat CPC1*col CPC2*col;
run;
proc mixed data = ANA;
  class row col treat;
  model Y = treat;
  random col CPC1*col CPC2*col;
  lsmeans treat;
run;

```

### AMMITA

/\*This program gives approximately the same results as the above AMMITA program but it is written solely for and AMMI analysis of columns and treatments.\*/

```

data colrow;
  infile 'colrow.dat';
  input Y row col treat;
proc glm data = colrow;
  class row col treat;
  model Y = col row treat;
  lsmeans col row treat;
proc sort;
  by row col;
proc means data = subres noprint;
  var resid;
  by row col;
  output out = resmean mean = Rmean;
proc sort;
  by row col;
proc transpose dat = resmean out = PCA;
  var Rmean;
  ID col;
  by row;
proc princomp cov data = PCA out = PCAout noprint;
  var _1 _2 _3 _4 _5 _6 _7; /*This is the number of columns, i.e., 7.*/
*proc print;
* by row;
* var prin1 prin2 prin3;
*run;
data crbig;
  set colrow;
  idx = _n_;
run;
proc sort;
  by col row;
run;
/*This part of the program generates a data set that contains the PCA and has the same number of
observations as the original data set crbig.*/
data pcaout1;
  set pcaout pcaout pcaout
    pcaout pcaout pcaout
    pcaout; /*"pcaout" has to be repeated the same number of times as there are columns, i.e., 7.*/
data crbig;
  merge crbig pcaout1;

```

```
    keep row col treat y idx prin1 prin2;
run;
proc sort data = crbig;
    by idx;
run;
proc print noobs; /*This is to verify that the new data set is correct.*/
run;
proc glm data = crbig;
    class row col treat;
    model Y = col treat prin1*col prin2*col;
run;
proc mixed data = crbig;
    class row col treat;
    model Y = treat;
    random col prin1*col prin2*col;
    lsmeans treat;
run;
```

## Appendix B. Programs for six analyses for balanced lattice square example.

/\*Let rep = replicate, col = column, and treat = treatment in the data file lsd1645.dat.\*/

### RCBD and textbook LSD

```
data lsd;
  infile 'lsd1645.dat';
  input count rep row col treat;
proc glm data = lsd;
  class rep row col treat;
  model y = rep treat;
  lsmeans treat; /*RCBD means*/
run;
proc glm data = lsd;
  class rep row col treat;
  model y = rep treat row(rep) col(rep);
  random rep row(rep) col(rep); /*For expected values of ANOVA mean squares*/
  lsmeans treat; /*Fixed effects means*/
run;
proc mixed data = lsd;
  class rep row col treat;
  model count = treat;
  random rep row(rep) col(rep);
  lsmeans treat; /*Treatment means with recovery of inter-row and inter-column information.*/
run;
```

### GRADTA and PRTA

```
proc iml;
  opn4 = orpol(1:4,3);
  opn4[,1] = (1:4)';
  op4 = opn4;
  create opn4 from opn4[colname = {'Col' 'C1' 'C2' 'C3'}];
  append from opn4;
  close opn4; run;
  opn3 = orpol(1:4,3);
  opn3[,1] = (1:4)';
  op3 = opn3;
  create opn3 from opn3[colname = {'ROW' 'R1' 'R2' 'R3'}];
  append from opn3;
  close opn3; run;
data lsd;
  infile 'lsd1645.dat';
  input count rep row col treat;
data lsbig; set lsd;
  idx = _n_; run;
proc sort data = lsbig;
  by COL; run;
data lsbig;
  merge lsbig opn4;
  by COL; run;
proc sort data = lsbig;
  by ROW; run;
data lsbig;
```

```

merge lsbig opn3;
by ROW; run;
proc sort data = lsbig;
by idx; run;
proc glm data = lsbig;
class rep row col treat;
model count = rep treat row(rep) c1*row(rep);
run;
proc mixed data = lsbig;
class rep row col treat;
model count = treat;
random rep row(rep) c1*row(rep);
lsmeans treat;
run;
proc glm data = lsbig;
class rep row col treat;
model count = rep treat c1*rep r1*rep r2*rep c1*r1*rep c2*r1*rep c2*r2*rep c3*r2*rep;
run;
proc mixed data = lsbig;
class rep row col treat;
model count = treat;
random rep c1*rep r1*rep r2*rep c1*r1*rep c2*r1*rep c2*r2*rep c3*r2*rep;
lsmeans treat;
run;

```

#### PCTA and AMMITA

```

data lsd;
infile lsd1645.dat;
input count rep row col treat;
label row = 'incomplete block';
proc sort data = lsd;
by rep row col;
proc iml;
use lsd;
read all var{rep row col treat} into class;
nrep = class[<>,1];
nro = class[<>,2];
nco = class[<>,3];
ntr = class[<>,4];
N = NROW(CLASS);
A2 = J(nrep*nro, nco + 2,0);
cont1 = 0;
rows = 1:nro;
do I = 1 to nrep;
read all where(rep = I) into A0;
A1 = shape(A0, nro, nco);
A2[(cont1 + 1:cont1 + nro),] = repeat(I,nro,1) || rows || A1;
cont1 = cont1 + nro;
end;
names0 = 'COL1': 'COL4'; names = {rep row} || names0;
create SPAT from A2 [colname = names];
append from A2;
quit;
proc print data = SPAT; run;

```

```

data colrow; set SPAT;
IF rep ne 1, THEN delete;
proc princomp data = colrow prefix = cpc out = colvar NOPRINT;      /*rep = 1*/
var COL1-COL4;
proc print data = colvar;
data t1; set col row; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW4;
data rowvar; set t2; col = _name_; drop _name_;
proc print data = rowvar;
  data c1; set colrow; count = cOL1; col = 'COL1'; keep row col count;
  data c2; set colrow; count = cOL2; col = 'COL2'; keep row col count;
  data c3; set colrow; count = cOL3; col = 'COL3'; keep row col count;
  data c4; set colrow; count = cOL4; col = 'COL4'; keep row col count;
data t3; set c1 c2 c3 c4;
proc sort; by row col;
run;
data CC; set colvar; drop COL1-COL4 CPC3-CPC4; proc print;
data RR; set rowvar; drop ROW1-ROW4 RPC3-RPC4; proc print;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO1; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
proc print;
data colrow; set SPAT;                                          /*rep = 2*/
IF rep ne 2, THEN delete;
proc princomp data = colrow prefix = cpc out = colvar NOPRINT;
var COL1-COL4;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW4;
data rowvar; set t2; col = _name_; drop _name_;
proc print data = rowvar;
  data c1; set colrow; count = cOL1; col = 'COL1'; keep row col count;
  data c2; set colrow; count = cOL2; col = 'COL2'; keep row col count;
  data c3; set colrow; count = cOL3; col = 'COL3'; keep row col count;
  data c4; set colrow; count = cOL4; col = 'COL4'; keep row col count;
data t3; set c1 c2 c3 c4;
proc sort; by row col;
run;
data CC; set colvar; drop COL1-COL4 CPC3-CPC4; proc print;
data RR; set rowvar; drop ROW1-ROW4 RPC3-RPC4; proc print;
proc sort data = t3; by row;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by ROW;
data TODO2; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
proc print;

```



```

data colrow; set SPAT;
IF rep ne 3, THEN delete;
proc princomp data = colrow prefix = cpc out = colvar NOPRINT;
var COL1-COL4;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW4;
data rowvar; set t2; col = _name_; drop _name_;
proc print data = rowvar;
  data c1; set colrow; count = cOL1; col = 'COL1'; keep row col count;
  data c2; set colrow; count = cOL2; col = 'COL2'; keep row col count;
  data c3; set colrow; count = cOL3; col = 'COL3'; keep row col count;
  data c4; set colrow; count = cOL4; col = 'COL4'; keep row col count;
data t3; set c1 c2 c3 c4;
proc sort; by row col;
run;
data CC; set colvar; drop COL1-COL4 CPC3-CPC4; proc print;
data RR; set rowvar; drop ROW1-ROW4 RPC3-RPC4; proc print;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO3; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
proc print;
data colrow; set SPAT;
IF rep ne 4, THEN delete;
proc princomp data = colrow prefix = cpc out = colvar NOPRINT;
var COL1-COL4;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW4;
data rowvar; set t2; col = _name_; drop _name_;
proc print data = rowvar;
  data c1; set colrow; count = cOL1; col = 'COL1'; keep row col count;
  data c2; set colrow; count = cOL2; col = 'COL2'; keep row col count;
  data c3; set colrow; count = cOL3; col = 'COL3'; keep row col count;
  data c4; set colrow; count = cOL4; col = 'COL4'; keep row col count;
data t3; set c1 c2 c3 c4;
proc sort; by row col;
run;
data CC; set colvar; drop COL1-COL4 CPC3-CPC4; proc print;
data RR; set rowvar; drop ROW1-ROW4 RPC3-RPC4; proc print;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO4; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
proc print;

```

```

data colrow; set SPAT;
IF rep ne 5, THEN delete;
proc princomp data = colrow prefix = cpc out = colvar NOPRINT;
var COL1-COL4;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW4;
data rowvar; set t2; col = _name_; drop _name_;
proc print data = rowvar;
  data c1; set colrow; count = cOL1; col = 'COL1'; keep row col count;
  data c2; set colrow; count = cOL2; col = 'COL2'; keep row col count;
  data c3; set colrow; count = cOL3; col = 'COL3'; keep row col count;
  data c4; set colrow; count = cOL4; col = 'COL4'; keep row col count;
data t3; set c1 c2 c3 c4;
proc sort; by row col;
run;
/*Creating one file (ANA) for analysis of the data.*/
data CC; set colvar; drop COL1-COL4 CPC3-CPC4; proc print;
data RR; set rowvar; drop ROW1-ROW4 RPC3-RPC4; proc print;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO5; set TODO0; COL = 1*COMPRESS(COL, 'COL');
proc sort; by ROW COL;
proc print;
data TODO; set TODO1 TODO2 TODO3 TODO4 TODO5;
proc print;
proc sort data = lsd out = t1; by REP ROW COL;
data t2; set t1; keep REP ROW COL TREAT;
data t3; set TODO; C = 1*COL; drop COL;
data t4; set t3; COL = C; drop C;
data ANA; merge t2 t4; by REP ROW COL;
/*ANOVAs and mixed model means*/
proc glm data = ANA;
  class rep treat;
  model count = rep treat CPC1(rep) CPC2(rep) RPC1(rep) CPC1*RPC1(rep)
    CPC1*RPC2(rep) CPC2*RPC1(rep) CPC2*RPC2(rep);
run;
proc mixed data = ANA;
  class rep treat;
  model count = treat;
  random rep CPC1(rep) CPC2(rep) RPC1(rep) CPC1*RPC1(rep)
    CPC1*RPC2(rep) CPC2*RPC1(rep) CPC2*RPC2(rep);
  lsmeans treat;
run;
proc glm data = ANA;
  class rep row col treat;
  model count = rep treat row(rep) RPC1*row(rep);
run;
proc mixed data = ANA;
  class rep row col treat;

```

```

model count = treat;
random rep row(rep) RPC1*row(rep);
lsmeans treat;
run;

```

#### AMMITA (For r replicates)

```

data blsd;
  infile 'blsd1645.dat'; input Y rep row col treat;
proc glm data = blsd;
  class rep row col treat;
  model Y = rep treat row(rep) col(rep);
  outhput out = subres R = resid;
proc sort;
  by row col;
proc means data = subres noprint;
  var resid;
  by row col;
output out = resmean out = PCA;
  var Rmean;
  ID col;
  by row;
proc princomp cov data = PCA out = PCAout noprint;
  var _1 _2 _3 _4; /*4 is the number of columns.*/
*proc print;
* by row:
* var prin1 prin2 prin3; /*The number of prin1 desired, 3 here.*/
*run;
data crbig;
  set blsd;
  idx = _n_;
run;
proc sort;
  by rep row col;
run;
data pcaout1;
  set pcaout pcaout pcaout pcaout; /*"pcaout" repeated c times, c = number of columns.*/
data pcaout2;
  set pcaout1 pcaout1 pcaout1 pcaout1; /*"pcaout1" repeated r times, r = number of replicates.*/
data crbig;
  merge crbig PCAout2;
  keep rep row col treat y idx prin1 prin2; /*prin3 not put in crbig.*/
run;
proc sort data = crbig;
  by idx;
run;
proc print noobs; /*This is to verify that new data set is what is desired.*/
run;
proc glm data = crbig;
  class rep row col treat;
  model Y = rep treat col(rep) prin2*rep; /*This is the final model.*/
run;
proc mixed data = crbig;
  class rep row col treat;
  model Y = treat;
  random rep col(rep) prin2*rep;
  lsmeans treat; run;

```

### Appendix C. Programs for six analyses for semi-balanced lattice square example

RCBD, SBLSD, PRTA, and GRADTA (These may be included in the same program.)

```
Proc iml;
  opn4 = orpol(1:5,4);
  opn4[,1] = (1:5)';
  op4 = opn4;
  create opn4 from opn4[colname = {'COL' 'C1' 'C2' 'C3' 'C4'}];
  append from opn4; close opn4;
run;
  opn3 = orpol(1:5,4);
  opn3[,1] = (1:5)'; op3 = opn3;
  create opn3 from opn3[colname = {'ROW' 'R1' 'R2' 'R3' 'R4'}];
  append from opn3; close opn3;
run;
data blsd;
  infile 'blsd2553.dat';
  inpute yield rep row col treat;
data lsbig; set blsd;
  idx = _n_; run;
proc sort data = lsbig;
  by COL; run;
data lsbig;
  merge lsbig opn4;
  by COL; run;
proc sort data = lsbig;
  by ROW; run;
data lsbig;
  merge lsbig opn3;
  by ROW; run;
proc sort dat = lsbig; by idx; run;
proc print; run;
proc glm data = lsbig;
  class rep row col treat;
  model yield = rep treat; run;
proc glm data = lsbig;
  class rep row col treat;
  model yield = rep treat row(rep) col(rep); run;
proc mixed data = lsbig;
  class rep row col treat;
  model yield = treat;
  random rep row(rep) col(rep);
  lsmeans treat; run;
proc glm data = lsbig;
  class rep row col treat;
  model yield = rep treat c1*rep r1*rep r2*rep r3*rep r4*rep c1*r1*rep c1*r2*rep; run;
proc mixed data lsbig;
  class rep row col treat;
  model yield = treat;
  random rep c1*rep r1*rep r2*rep r3*rep r4*rep c1*r1*rep c1*r2*rep;
  lsmeans treat; run;
proc glm data = lsbig;
  class rep row col treat;
```

```

model yield = rep treat row(rep) c1*row(rep); run;
proc mixed data lsbig;
class rep row col treat;
model yield = treat;
random rep row(rep) c1*row(rep);
lsmeans treat; run;

```

#### PCTA and AMMITA

```

data sbis;
infile 'blsd2553.dat'; input Y rep row col treat;
label rows = 'incomplete block';
proc sort data = blsd; by REP ROW COL;
proc iml;
use blsd;
read all var{REP ROW COL TREAT} into class;
NREP= class[<>,1];
NRO = class[<>,2];
NCO= class[<>,3];
NTR = class[<>,4];
N = NROW(class);
A2 = J(NREP*NRO,NCO+2,0);
CONT1 = 0;
ROWS = 1:NRO;
DO I = 1 TO NREP;
READ ALL WHERE(REP=I) VAR(Y) INTO A0;
A1 = SHAPE(A0, NRP, NCO);
A2[(CONT1+1:CONT1+NRO),] = REPEAT(I,NRO,1) || ROWS` || A1;
CONT1 = CONT1 + NRO;
END;
NAMES0 = 'COL1':'COL5'; NAMES = (REP ROW) || NAMES0;
CREATE SPAT FROM A2 [COLNAME = NAMES];
APPEND FROM A2; QUIT;
PROC PRINT DATA = SPAT; RUN;
data colrow; set SPAT; if REP NE 1 THEN DELETE; /*rep = 1.*/
proc princomp data = blsd prefix = cpc out = colvar NOPRINT;
var COL1-COL5;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW5;
data rowvar; set t2; col = _name_; drop _name_ ;
proc print data = rowvar;
data c1;set colrow; y = cOL1; col = 'COL1'; keep row col y;
data c2;set colrow; y = cOL2; col = 'COL2'; keep row col y;
data c3;set colrow; y = cOL3; col = 'COL3'; keep row col y;
data c4;set colrow; y = cOL4; col = 'COL4'; keep row col y;
data c5;set colrow; y = cOL5; col = 'COL5'; keep row col y;
data t3; set c1 c2 c3 c4 c5; /*Number of columns.*/
proc sort; by row col; run;
data CC; set colvar; drop COL1-COL5 CPC3-CPC5;
data RR; set rowvar; drop ROW1-ROW5 RPC3-RPC5;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;

```

```

proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO1; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
data colrow; set SPAT; if REP NE 2 THEN DELETE; /*rep = 2.*/
proc princomp data = blsd prefix = cpc out = colvar NOPRINT;
var COL1-COL5;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW5;
data rowvar; set t2; col = _name_; drop _name_ ;
proc print data = rowvar;
  data c1;set colrow; y = cOL1; col = 'COL1'; keep row col y;
  data c2;set colrow; y = cOL2; col = 'COL2'; keep row col y;
  data c3;set colrow; y = cOL3; col = 'COL3'; keep row col y;
  data c4;set colrow; y = cOL4; col = 'COL4'; keep row col y;
  data c5;set colrow; y = cOL5; col = 'COL5'; keep row col y;
  data t3; set c1 c2 c3 c4 c5; /*Number of columns.*/
proc sort; by row col; run;
data CC; set colvar; drop COL1-COL5 CPC3-CPC5;
data RR; set rowvar; drop ROW1-ROW5 RPC3-RPC5;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO1; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
data colrow; set SPAT; if REP NE 3 THEN DELETE; /*rep = 3.*/
proc princomp data = blsd prefix = cpc out = colvar NOPRINT;
var COL1-COL5;
proc print data = colvar;
data t1; set colrow; drop REP ROW;
proc transpose data = t1 out = rowcol prefix = ROW;
proc princomp data = rowcol prefix = rpc out = t2 NOPRINT;
var ROW1-ROW5;
data rowvar; set t2; col = _name_; drop _name_ ;
proc print data = rowvar;
  data c1;set colrow; y = cOL1; col = 'COL1'; keep row col y;
  data c2;set colrow; y = cOL2; col = 'COL2'; keep row col y;
  data c3;set colrow; y = cOL3; col = 'COL3'; keep row col y;
  data c4;set colrow; y = cOL4; col = 'COL4'; keep row col y;
  data c5;set colrow; y = cOL5; col = 'COL5'; keep row col y;
  data t3; set c1 c2 c3 c4 c5; /*Number of columns.*/
proc sort; by row col; run;
data CC; set colvar; drop COL1-COL5 CPC3-CPC5;
data RR; set rowvar; drop ROW1-ROW5 RPC3-RPC5;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO1; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;

```

```

data CC; set colvar; drop COL1-COL5 CPC3-CPC5;
data RR; set rowvar; drop ROW1-ROW5 RPC3-RPC5;
proc sort data = t3; by ROW;
data t6; merge t3 CC; by ROW;
proc sort data = t6; by COL;
data TODO0; merge t6 RR; by COL;
data TODO5; set TODO0; COL = 1*COMPRESS(COL,'COL');
proc sort; by ROW COL;
data TODO; set TODO1 TODO2 TODO3; /*Number of reps.*/
proc sort data blsd out = t1; by REP ROW COL;
data t2; set t1; keep REP ROW COL TREAT;
data t3; set TODO; C + 1*COL; drop COL;
data t4; set t3; COL = C; drop C;
data ANA; merge t2 t4; by REP ROW COL;
proc print;
proc glm data = ANA;
  class REP ROW COL TREAT;
  model Y = rep treat CPC1*rep CPC2*rep RPC1*rep RPC2*rep CPC1*RPC1*rep
CPC1*RPC2*rep
  CPC2*RPC1*rep CPC2*RPC2*rep; run;
proc mixed data = ANA;
  class REP ROW COL TREAT;
  model Y = treat;
  random rep CPC1*rep CPC2*rep RPC1*rep CPC1*RPC1*rep RPC2*rep CPC1*RPC2*rep
CPC2*RPC1*rep CPC2*RPC2*rep;
  lsmeans treat; run;
proc glm data = ANA;
  class REP ROW COL TREAT;
  model Y = rep treat row(rep) RPC1*row(rep); run;
proc mixed data = ANA;
  class REP ROW COL TREAT;
  model Y = treat;
  random rep row(rep) RPC1*row(rep);
  lsmeans treat; run;

```

### AMMITA

(Use code for balanced lattice square example; change number of columns and number of replicate. Also, will need to interchange row and column designation in input data as there are larger differences among rows than there is among columns for this example.)